UNIVERSITY OF COPENHAGEN

Konvertering af filer fra BySoc til DGCSS-CHAT

Beskrivelse af omsætningsprocesser, notationer og filformater

Lind, Peter; Jensen, Torben Juel

Publication date:
2006

Document version
Også kaldet Forlagets PDF

Citation for published version (APA):
Lind, P., & Jensen, T. J. (2006). *Konvertering af filer fra BySoc til DGCSS-CHAT: Beskrivelse af omsætningsprocesser, notationer og filformater*. (1 udg.) (s. 1-14). Danmarks Grundforskningsfonds Center for Sociolingvistiske Sprogforandringsstudier. http://dgcss.hum.ku.dk/upload/application/pdf/f51d6748/BySoc-beskrivelse.pdf

Download date: 05. May. 2023

Konvertering af filer fra BySoc til DGCSS-CHAT

Beskrivelse af omsætningsprocesser, notationer og filformater

af
Peter Lind og Torben Juel Jensen

Resume:
Dette dokument er en del af dokumentationen til BySocImport – et stykke software til automatisk konvertering af transskriptioner fra BySoc-projektet til et format brugbart af DGCSS-projektet. Denne del af dokumentationen er mestendels en oversigt over koder, symboler og datafelter. Den er forsøgt holdt så lidt teknisk som muligt, og er primært tænkt som en dokumentation af de ændringer af BySoc-transskriptionerne der er sket for at bringe dem i overensstemmelse med DGCSS' udskrivningskonvention.

Introduktion

I forbindelse med projekt By-Sociolingvistik, kaldet BySoc, er der transskriberet et antal interviews og gruppesamtaler. Disse transskriptioner ønsker man at analysere i DGCSS-projektet, men de er skrevet i et andet udskrivnings- og filformat end det der er valgt i her, og skal derfor konverteres. Dette dokument beskriver de to formater i forhold til hinanden, med særlig henblik på de specielle tegn og symboler hvert format benytter. Selve konverteringsprocessen er sket ved hjælp af programmet BySocImport, som er en del af DGCSS Suite. De tekniske specifikationer for dette system, som er udarbejdet af Peter Lind, beskrives i rapporten ”DGCSS Suite”.

Filformater

BySoc er transskriberet efter Dansk Standard 2 (Gregersen 1991), og oprindelig holdt i filformatet KUATEKST, men filerne har efterfølgende været igennem en ’renselsesproces’, primært foretaget af Peter Juel Henrichsen, og er derfor i et ganske særegent filformat, hvor hver talers talestrøm er i en fil for sig, og hvor fænomener som pauser og tøven ikke er markeret med bogstaver, men med symboler. Formatet er ganske velbeskrevet i ”Transliteration between spoken language corpora” (Allwood et al 2005, p.6 pp).

DGCSS transskriberer i en modificeret udgave af CHAT-formatet (MacWhinney). CHAT er et af de to formater som programmet CLAN kan arbejde med. Formatet er beskrevet i DGCSS’ udskrivningsmanual. Formatet kaldes her i dokumentet DGCSS-CHAT.

Under arbejdet med konverteringen har det vist sig fordelagtigt at indlemme et tredje format. Da BySoc filerne indeholder informationer der ikke kan noteres i DGCSS-formatet, og DGCSS-formatet samtidig kan indeholde informationer der ikke er noteret i BySoc, vil en konvertering fra det ene til det andet altid medføre et informationstab. For at udskyde dette informationstab til så sent i processen som muligt er der oprettet et superformat der inkluderer alle de informationer begge formater kan registrere. Superformatet eksisterer fortrinsvist internt i computerens hukommelse under konverteringen, men det er muligt at gemme det i form af en XML-fil. Dette dokument er en sammenskrivning af specifikationerne for to programmer, et der konverterer fra BySoc til superformatet, og et der konverterer fra superformatet til DGCSS-CHAT.

Hoved og krop

Ud over selve transskriptionen indeholder en fil, uanset format, en del ekstra-informationer. Det kan for eksempel være informationer om samtaleformen, tidspunktet for transskriptionen, informationer om deltagerne i samtalen, med mere. Denne ekstra-information kaldes også header-information fordi den ofte placeres i starten af filen, altså i fil-hovedet.
I forbindelse med konverteringen er der enkelte informationer der skal viderebringes fra BySoc-formatet, og da disse informationer er adskilt fra selve transskriptionen, behandles de også særskilt her i dokumentationen.

Der er mange header-informationer i BySoc-formatet, mange flere end der kan overføres til DGCSS-CHAT. Men de fleste informationer, for eksempel om deltagerne, er allerede gemt i en database, og DGCSS-CHAT-filen skal derfor blot henvise til deltagere med en unik kode hvormed de kan identificeres i databasen.

Partitur og taleture

Den største forskel på BySoc-formatet og DGCSS-CHAT er måden hvorpå de enkelte deltageres tale er opdelt. Begge formater benytter sig af opdeling i ortografisk noterede ord, men i BySoc har hver taler sin egen linje, der strækker sig gennem hele transskriptionen – et såkaldt partiturformat – hvor DGCSS-CHAT opdeler talen i individuelle ture.

Et kort udsnit af de to formater viser tydeligt forskellen:

BySoc udsnit

```
1: ja det var vist en af mine klassekammerater £                                    ja hvis
2:                                          nå var det en af dine klassekammerater £    nå
```

Og det samme tekstudsnit i DGCSS-CHAT:

```
*KKJ:    ja det var vist en af mine klassekammerater #.
*XHH:    nå var det en af dine klassekammerater #.
*KKJ:    ja <hvis> [>] [?] det var hende der Rie [?] så er det
         er en af mine klassekammerater.
*XHH:    <nå> [<].
```

BySoc-formatet viser den ene taler i linje 1 og den anden i linje 2. Disse to linjer fortsætter så ud mod højre i hele transskriptionens varighed. DGCSS-formatet har opdelt talen i 'ture': Først siger den ene taler noget, så den anden, så den ene igen og så fremdeles.

Konverteringsprogrammet skal så læse 'talestrømmen' i BySoc-formatet og klippe den op i individuelle "taleture". En opdeling i taleture kan af og til forekomme lidt kunstig og er under alle omstændigheder baseret på en fortolkning af samtalesituationen, mens der i forbindelse med konverteringsprocessen vil der være tale om en særdeles grov og maskinel opdeling. Så snart en deltager har et ophør i sin tale – et ophør der i den grafiske fremstilling er længere end mellemrummet mellem to ord – regnes turen for afsluttet.

Overlap

I BySocs partiturformat er overlappende tale ikke angivet eksplicit. Det er op til læseren at vurdere om to talere siger noget samtidigt. For nu at blive i musik-analogien så kan man antage at der er et taktslag for hvert tegn i teksten (og for hvert mellemrum i teksten for den tavse deltager). Hvis der på et givent taktslag er noget tekst ved både taler1 og taler2, så må de tale samtidig, og den ene overlapper følgelig den anden.

Den taler der havde turen (og bliver overlappet), kaldes 'den overlappede', og den taler der overlapper, kaldes 'den overlappende'. I det ovenstående eksempel overlapper taler2's "nå" taler1's "hvis".

I DGCSS-CHAT markeres både den overlappede og den overlappende tekst med `<` før det første og `>` efter det sidste ord i overlappet. Efter den afsluttende `>` angives overlappets type (overlappet eller overlappende), og hvis der er flere overlap i en taletur, nummereres hvert enkelt. Hvis taleren er den overlappede, angives det med `[>]`, og hvis taleren er den overlappende, angives det med `[<]`.

Konvertering

Konverteringen er som nævnt todelt: Først indlæses BySoc-filen til et superformat, derefter skrives superformatet til en DGCSS-CHAT-fil. Mellem disse to trin kan man for eksempel ændre i header-informationer, herunder indsætte deltagerkoder.

Header-informationer

DGCSS-CHAT-filhovedet ser således ud:

```
@Begin
@Languages: da
@Participants: kode navn rolle, kode navn rolle
@ID: dgcss
@Situation: situation
@Comment: interviewddmmåå, København, interview-id kommentarer
@Transcriber: transskribtør
@Checker:
@Transcription date:
@File: filnavn
```

De kursiverede tekster skal udfyldes under konverteringen. Det mest interessante er `@Participants`-linjen, der skal udfyldes med en liste over alle deltagerne i transskriptionen. Hver deltager har en trebogstavskode der er unik i DGCSS-projektet, et navn der ikke må indeholde mellemrum, og en rolle der skal være enten `Interviewer` eller `Meddeler` (som skrevet, med stort begyndelsesbogstav). Hver deltager adskilles af et komma.

Trebogstavskoden blev ikke benyttet i BySoc-projektet og kan derfor ikke findes automatisk. Den skal indsættes manuelt under konverteringen – i et særligt tekstfelt i programmet.

I BySoc-formatet er alle header-informationer for samtlige samtaler og deres transskriptioner samlet i én fil kaldet `extralin.txt`. Denne fil har informationerne opdelt i blokke. Der er en blok for hver samtale, og denne blok indeholder udover informationer om samtalen, yderligere blokke. Der er en blok for hver deltager, og en blok for hver transskription der måtte være af den pågældende samtale. Når der er flere forskellige transskriptioner, er det ofte uddrag til stilanalyse eller lignende, og de skal som regel ikke konverteres da der ikke vil blive arbejdet videre på dem.

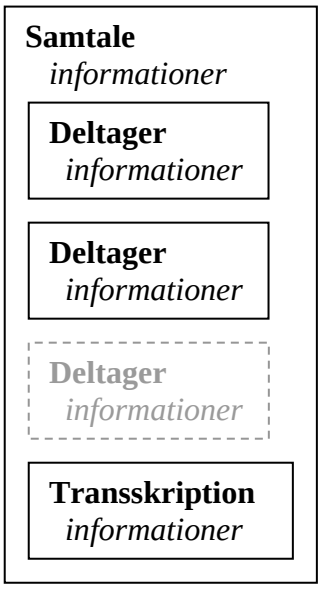

Diagram over blokke i extralin.txt

De tre blokke gengives i tre tabeller herunder. De angives med en beskrivelse af indholdet, BySoc-navnet (som regel en firebogstavskode) og den linje i DGCSS-CHAT-filhovedet hvori data indsættes. Når et datafelt kun kan indeholde to forskellige faste værdier, er de angivet direkte under feltet, med ekstra indrykning og uden beskrivelse.

Når et datafelt ikke medtages i DGCSS-CHAT-filen, er det markeret med grå baggrund.

Samtale

Betegnelse	**BySoc**	**DGCSS-CHAT**
Interview ID	INTERVIEW	`@Comment:` *interview-id*
Recording ID (bånd nr.)	BDNR	
Filepath	BS96	
Varighed i minutter	ITLE	
Antal deltagere	ADEL	
Antal transskriptioner	ATRS	
Type	BSTY	`@Situation:`
	pers	`interview`
	gr	`gruppesamtale`
Kommentarer	EVTI	

Deltagere

Betegnelse	**BySoc**	**DGCSS-CHAT**
Deltagerkode	DELTAGER	
BySoc ID	BSID	
Gruppe/kategori	BSGR	
Rolle	ROLL	*rolle:*
	itv	`Interviewer`
	inf	`Meddeler`
Navn	NAVN	*navn*
Initialer	INIT	
Alder	ALDR	
Køn	KOEN	
	M	
	K	
Socialklasse	KLAS	
	AK	
	MK	
Tilhørsforhold, dialektalt	TILH	
Kommentarer	EVTD	

Transskriptioner

Betegnelse	**BySoc**	**DGCSS-CHAT**
Transskriptions ID	TRANSSKRIPTION	
Originalfilens placering	BS97	
Dækning	TRDK	
	T	
	E	
Varighed	ITTR	
Transkriptør	TRAN	`@Transcriber:`
Kommentarer	EVTT	`@Comment:` *kommentarer*

Transskriberet tale

Når header-informationerne er konverteret, kommer turen til filens krop, selve transskriptionen. Som tidligere beskrevet klippes talestrømmen op i individuelle ture, og overlap markeres. Derudover sker der en konvertering af tegn og symboler benyttet i transskriptionen.

Ord

Ortografien overføres næsten uændret. I BySoc er der dog i visse tilfælde benyttet udtaletilnærmelse i form af udeladte bogstaver markeret med en apostrof hvilket ikke er tilladt i DGCSS-CHAT. Det drejer sig for eksempel om ord som ”ikke” der er noteret ik’. Dette og andre ord med apostrof oversættes under konverteringen efter følgende tabel.

BySoc	DGCSS-CHAT
altså'	altså
alså'	altså
BZ'ere	bz'ere
Citröen'en	Citröenen
eller'	eller
gu'	gud
Honda'en	Hondaen
hva'	hvad
hvad'	hvad
hva'bar'	hvadbehager
hva'behar'	hvadbehager
hvo'n	hvordan
ik'	ikke
ik''	ikke
ik'altså	ikke altså
ik'alså'	ikke altså
ik'også	ikke også
ik'også'	ikke også
ik'og	ikke også
ik'sandt'	ikke sandt
ik'så	ikke så
joik'	jo ikke
Lada'en	Ladaen
LP'er	lp'er
PC'er	pc'er
sgu'	sgu
så'	så
sårn'	sådan
vel'	vel
østerbro'ere	østerbroere

Forkortelser

Hvis der benyttes forkortelser som eksempelvis `cand.mag.` i det transskriberede, bliver punktummerne fjernet. Konverteringen fjerner konsekvent alle punktummer i teksten.

Bogstavforkortelser som `EFG`, bliver ændret til `E_F_G`. Udtalelige akronymer som `Nato` bevares som de er skrevet. Konverteringsprogrammet kigger udelukkende på forekomsten af versaler, så hvis der for eksempel står `NATO`, vil det blive ændret til `N_A_T_O`.

Ikke-komplette ord

Hvis en deltager ikke siger et ord færdigt, men afbryder sig selv eller bliver afbrudt af en anden deltager, skal ordet markeres som afbrudt. BySoc-formatet markerer afbrydelser med en bindestreg før eller efter det afbrudte ord – bindestregen repræsenterer den manglende del af ordet – men skelner ikke som DGCSS-CHAT mellem afbrydelser af enkelte ord i en taletur og afbrydelser der fører til turskift. Da de to typer afbrydelser er markeret ens i BySoc-formatet, sørger konverteringsprogrammet for at ændre en selvafbrydelse til en tur-afbrydelse, hvis det afbrudte ord er det sidste i turen.

Betegnelse	**BySoc**	**DGCSS-CHAT**
Selvafbrydelse	*(det afbrudte ord)*-	*(det afbrudte ord)* `[/]`
Genoptagelse af afbrudt ord	-*(det afbrudte ord)*	`§`*(det afbrudte ord)*
Tur-afbrydelse	*(det afbrudte ord)*-	*(det afbrudte ord)* `+/.`

Kommentarer

I BySoc-formatet er de fleste kommentarer holdt i et særskilt kommentarspor, altså en tekstlinje ud over de deltagendes. Nogle gange er kommentarer dog angivet i parenteser i en udvalgt deltagers tale. Det forekommer som regel når en taler hoster, snøfter eller foretager en anden ikke-sproglig handling. I de følgende afsnit beskrives de parentes-kommentarer der oversættes til henholdsvis latter, vejrtrækning og ikke-tale. Øvrige parentes-kommentarer bliver behandlet som almindelige kommentarer, altså dem der er holdt i kommentarsporet.

I DGCSS-CHAT bliver alle kommentarer anbragt i en `%com`-linje efter den relevante taletur. For kommentarer i talestrømmen (altså de ovennævnte parentes-kommentarer) bliver det foregående ord også skrevet i `%com`-linjen. Ordet skrives i anførselstegn, sådan: `%com: "ord": kommentar`

Latter

Latter noteres i DGCSS-CHAT altid som `ha`. Der skelnes ikke mellem forskellige former, længde eller vokalkvalitet i latteren.

I BySoc-formatet er der benyttet flere forskellige notationer for latter, nogle gange i form af en beskrivende parentes-kommentar, og andre gange i talestrømmen, ofte med varians på vokalkvaliteten (`ha`, `hi`, `hø`, `he`). Konverteringen ændrer ikke på latter noteret i talestrømmen, men følgende parentes-kommentarer konverteres til `ha`:

fnis	hi hi	ler højlydt
fnis?	hysterisk latter	ler kort
fnis~	hø hø	ler lidt
fniser	høj latter	ler stadig
forlegent grin	kluk	ler?
genert latter	klukker	let genert latter
genert leende	klukler	let latter
grin	latter	let leende
griner	latter, leende	lille latter
ha	latter,hø	markeret latter
ha ha	leen	markeret leende
ha ha ha	ler	smågrin
haha	ler ler	smålatter
hi	ler genert	småler

Vejrtrækning

I DGCSS-CHAT noteres vejrtrækning som `hh`. Der skelnes ikke mellem ind- og udånding. BySoc benytter symbolet # for en pause udfyldt med vejrtrækning, som umiddelbart kan konverteres til `hh`, men af og til er der også en markering af vejrtrækningen i en parentes-kommentar.

Følgende parentes-kommentar konverteres til `hh`:

trækker vejret

Ikke-tale

Lyde, hvad enten de er fremkaldt af samtalens deltagere eller er baggrundsstøj, markeres i DGCSS-CHAT med `[% ... ]` hvor ... er en beskrivelse af lyden.
I BySoc er der ikke nogen standard for notation af lyde, men ganske ofte er de markeret som parentes-kommentarer ved den taler der frembringer lyden.

Følgende liste opremser de kommentarer der konverteres til en lyd-angivelse.

BySoc	**DGCSS-CHAT**
banken	[% banken]
banker i bordet	[% banker i bordet]
barnet pludrer	[% barnet pludrer]
bladrer i papirer	[% bladrer i papirer]
bladrer	[% bladrer]
brummer	[% brummer]
det banker	[% det banker]
drikker kaffe	[% drikker kaffe]
dybt suk	[% dybt suk]
dør smækker	[% dør smækker]
fløjt	[% fløjt]
fløjtelyd	[% fløjtelyd]
fløjter ud gennem den ene side af munden	[% fløjter ud gennem den ene side af munden]
fløjter	[% fløjter]
fnys	[% fnys]
fnyser	[% fnyser]
gaber	[% gaber]
giver lyd	[% giver lyd]
giver sig	[% giver sig]
græder	[% græder]
host	[% host]
host host	[% host]
hosten	[% host]
hoster	[% host]
hvislelyd	[% hvislelyd]
hyler	[% hyler]
hælder kaffe op	[% hælder kaffe op]
irritationslyd	[% irritationslyd]
klapper	[% klapper]
klir af kaffekopper	[% klir af kaffekopper]
klir af kop	[% klir af kop]
klir	[% klir]

BySoc	DGCSS-CHAT
klynker	[% klynker]
knas	[% knas]
knipser med fingrene	[% knipser med fingrene]
knipser	[% knipser]
larmer meget	[% larmer meget]
laver illustrerende lyd	[% laver illustrerende lyd]
laver lyd	[% laver lyd]
laver lyde	[% laver lyde]
laver rallelyde	[% laver rallelyde]
lyd	[% lyd]
læbelyd	[% læbelyd]
lægger røret på	[% lægger røret på]
lægger røret	[% lægger røret]
nynner	[% nynner]
nyser	[% nyser]
pludrer	[% pludrer]
pust	[% pust]
puster røg ud	[% puster røg ud]
puster ud	[% puster ud]
puster	[% puster]
ryger	[% ryger]
rømmen	[% rømmen]
rømemr sig	[% rømmer sig]
rømme sig	[% rømmer sig]
rømmer sig	[% rømmer sig]
råber og skriger	[% råber og skriger]
skriger	[% skriger]
skænker kaffe	[% skænker kaffe]
smæk med læber	[% smæk med læber]
snork	[% snork]
snøft	[% snøft]
snøfter	[% snøft]
spyttelyd	[% spyttelyd]
stiller kop	[% stiller kop]
stiller noget på bordet	[% stiller noget på bordet]
støj	[% støj]
støn	[% støn]
stønner	[% stønner]
suk	[% suk]
sukker	[% suk]
synker	[% synker]
telefonen ringer	[% telefonen ringer]
tsk	[% tsk]
tsk~	[% tsk]
tænder cigaret	[% tænder cigaret]
tænder en smøg	[% tænder en smøg]
uro	[% uro]
vrængende lyd	[% vrængende lyd]

Transskriptionsnoter

Når transskriptøren ikke med sikkerhed har kunnet transskribere en passage, kan den enten være markeret som tvivlsom eller uforståelig. En tvivlsom passage er forsøgt transskriberet, men omkranset med koder der afskærer den fra den omkringliggende tekst. En uforståelig passage er ikke forsøgt transskriberet, men erstattet af en kode.

Betegnelse	**BySoc**	**DGCSS-CHAT**
Tvivlsom passage	[…]	[?] … [?]
Uforståelig passage	(uf) [1)]	xxx

1) Den officielle standard for uforståelig tale i BySoc er (uf), men der forekommer også andre angivelser, blandt andet tastefejlen (fu) og den uddybende (uforståeligt). Konverteringen oversætter både uf, fu og alle parentes-kommentarer der begynder med bogstaverne uf, til xxx.

Meningsbærende ikke-tale

Lyde der er frembragt af taleorganerne, men ikke kan karakteriseres som tale, bliver for det meste angivet på samme måde som andre lyde, blandt andet baggrundsstøj (jf. ovenfor). BySoc markerer hørbar ånding og tøven med særlige notationer. Af og til er andre lydelige fænomener noteret i form af ikke standardiserede onomatopoietika, som for eksempel ”pfzi” der efter alt at dømme angiver et hvislende pust. Det er ikke muligt at konvertere sidstnævnte automatisk, så de må rettes manuelt under korrekturlæsningen.

Betegnelse	**BySoc**	**DGCSS-CHAT**
Pause med ånding	#	hh [1)]
Tøven	~ [2)]	øh
Tyssen		sh
Returytring		mm
Udbrud		uh
Latter	(latter) [3)]	ha

1) ”Transliteration between spoken corpora” kalder pause med ånding for ”exhalation”,, men der skelnes ikke mellem ind- og udånding i konverteringen
2) I BySoc benyttes tegnet ~ både som angivelse af en fyldt pause, altså en tøven, og som angivelse af vokalforlængelse i et ord, altså en tøvende udtale. Men der angives ikke hvilken vokal der er forlænget, så det kan ikke umiddelbart oversættes til DGCSS-CHATs markering af vokalforlængelse. Et ~ bliver altså altid konverteret til et øh, uanset om det er hæftet til et ord, eller står alene.
3) Se listen på side 8 for en komplet oversigt over de latter-angivelser der konverteres.

Pauser

Der opereres med tre forskellige pauselængder i BySoc, angivet med henholdsvis et, to eller tre pausesymboler. I DGCSS-CHAT angives en pause altid kun med et symbol, uanset varigheden. En undtagelse er de steder hvor taler2 overlapper taler1 imellem to ord, uden at taler1 holder en pause, og uden at taler2 overlapper taler1's ord. Der skelnes således mellem pauser hvor ingen af talerne siger noget, og "pauser" hvor én taler har et kort ophold i talen som samtalepartneren udfylder med en returytring. Da BySocs partiturformat ikke muliggør at en taler kan sige et helt ord (i det mindste ikke et ord på mere end et bogstav) imellem to af en anden talers ord, vil dette fænomen ikke kunne være noteret, og det ignoreres derfor i konverteringen.

Betegnelse	**BySoc**	**DGCSS-CHAT**
Pause	£ ££ £££	#
overlappet mellemrum		€

Udtalemodifikationer

Betegnelse	**BySoc**	**DGCSS-CHAT**
Emfase		[!!]
Citeret tale	" … "	
Stigende intonation	? [1]	
Vokalforlængelse	~ [2]	:

1) I BySoc er ? ofte brugt efter spørgsmål hvor en del af de foregående ord givetvis har haft en stigende intonation. I DGCSS-CHAT noteres intonationsændringer ikke, men det er heller ikke muligt at konvertere symbolet da det ikke angiver hvornår den stigende intonation begynder, kun hvornår den slutter.

2) Vokalforlængelse angives i DGCSS-CHAT med et kolon efter den forlængede vokal. I BySoc med en tilde før eller efter det ord der har en forlænget vokal. Da det samme symbol også bruges til tøvende udtale, kan konverteringen ikke oversætte til en vokalforlængelse og benytter i stedet strategien beskrevet i afsnittet: Meningsbærende ikke-tale pkt. 2)

Nummerering af overlap

Overlappende og overlappet tale markeres i DGCSS-CHAT med ét symbol før og et andet symbol efter både den overlappede og den overlappende passage. Efter passagen angives endvidere om den var overlappet eller overlappende. Hvis der er flere overlap-passager i samme taletur, nummereres de fra 1 og op. Hvis der kun er en overlap-passage, har den ikke noget nummer. Den overlappede og den overlappende passage skal have samme nummer.

Betegnelse	**BySoc**	**DGCSS-CHAT**
Overlap start (for den overlappede)	*Endnu en taler begynder at tale.*	<
Overlap start (for den overlappende)	*En anden taler var i gang med at tale*	<
Overlap slut (for den overlappede)	*Antallet af samtidige talere ændres*	> [>*N*] [1]
Overlap slut (for den overlappende)	*Antallet af samtidige talere ændres*	> [<*N*] [1]

1) *N* er et tal fra 1 og opefter, eller ikke angivet.

Kodeoversigt

Betegnelse	**BySoc**	**DGCSS-CHAT**
Tvivlsom passage	[…]	[?] … [?]
Uforståelig passage	(uf)	xxx

Betegnelse	**BySoc**	**DGCSS-CHAT**
Pause	£ ££ £££	#
Overlappet mellemrum		€

Betegnelse	**BySoc**	**DGCSS-CHAT**
Pause med ånding	#	hh
Tøven	~	øh
Tyssen		sh
Returytring		mm
Udbrud		uh
Latter	(latter)	ha

Betegnelse	**BySoc**	**DGCSS-CHAT**
Selvafbrydelse	-	[/]
Genoptagelse af afbrudt ord	-	§
Afbrydelse (i forbindelse med turskift)	-	+/.

Betegnelse	**BySoc**	**DGCSS-CHAT**
Emfase		[!!]
Citeret tale	” … ”	
Stigende intonation	?	
Vokalforlængelse *(bliver dog ikke konverteret – se side 11)*	~	:

Betegnelse	**BySoc**	**DGCSS-CHAT**
Overlap start (for den overlappede)	*Endnu en taler begynder at tale.*	<
Overlap start (for den overlappende)	*En anden taler var i gang med at tale*	<
Overlap slut (for den overlappede)	*Antallet af samtidige talere ændres*	> [>*N*]
Overlap slut (for den overlappende)	*Antallet af samtidige talere ændres*	> [<*N*]

Referencer

Allwood, Jens et al: ”Transliteration between spoken language corpora” i Nordic Journal of Linguistics 28.1, side 1-32, 2005 Cambridge University Press

Gregersen, Kirsten: “Dansk Standard for udskrifter og registrering af talesprog” 2. udgave, 1992, Institut for Sprog og Kommunikation, Odense Universitet.

MacWhinney, Brian: Child Language Data Exchange System http://childes.psy.cmu.edu

Pharao, Nicolai et al: ”Udskrivningsmanual for DGCSS, februar 2006”, DGCSS. Tilgængelig via DGCSS’ hjemmeside. http://dgcss.hum.ku.dk/organisation/praktiske_oplysninger/